

Product Superposition for MIMO Broadcast Channels

Yang Li, *Student Member, IEEE*, and Aria Nosratinia, *Fellow, IEEE*

Abstract—This paper considers the multiantenna broadcast channel without transmit-side channel state information (CSIT). For this channel, it has been known that when all receivers have channel state information (CSIR), the degrees of freedom (DoF) cannot be improved beyond what is available via TDMA. The same is true if none of the receivers possess CSIR. This paper shows that an entirely new scenario emerges when receivers have unequal CSIR. In particular, orthogonal transmission is no longer DoF-optimal when one receiver has CSIR and the other does not. A multiplicative superposition is proposed for this scenario and shown to attain the optimal degrees of freedom under a wide set of antenna configurations and coherence lengths. Two signaling schemes are constructed based on the multiplicative superposition. In the first method, the messages of the two receivers are carried in the row and column spaces of a matrix, respectively. This method works better than orthogonal transmission while reception at each receiver is still interference-free. The second method uses coherent signaling for the receiver with CSIR, and Grassmannian signaling for the receiver without CSIR. This second method requires interference cancellation at the receiver with CSIR, but achieves higher DoF than the first method.

I. INTRODUCTION

In the MIMO broadcast channel, when channel state information is available at the receiver (CSIR) but not at the transmitter (CSIT), orthogonal transmission (e.g., TDMA) achieves optimal degrees of freedom (DoF) [1], [2]. With neither CSIT nor CSIR, again orthogonal transmission achieves the best possible DoF [3]. This paper studies the broadcast channel where one receiver has full CSIR and another has no CSIR. In this case, new DoF gains are discovered that can be unlocked with novel signaling strategies.

The study of broadcast channels with unequal CSIR is motivated by downlink scenarios where users have different mobilities. Subject to a fixed downlink pilot transmission schedule, the low-mobility users have the opportunity to reliably estimate their channels, while the high-mobility users may not have the same opportunity.

The main result of this paper is that when one receiver has full CSIR and the other has none, the achieved DoF is strictly better than that obtained by orthogonal transmission. For the unequal CSIR scenario, we propose a *product superposition*, where the signals of the two receivers are multiplied to produce the broadcast signal. In the following the receiver with full

CSIR is referred to as the *static receiver* and the receiver with no CSIR as the *dynamic receiver*. Two classes of product superposition signaling are proposed:

- In the first method, information for both receivers is conveyed by the row and column spaces of a transmit signal matrix, respectively. The signal matrix is constructed from a product of two signals that lie on different Grassmannians. The two receivers do *not* interfere with each other even though there is no CSIT, a main point of departure from traditional superposition broadcasting [1], [4].
- In the second method, information for the static receiver is carried by the signal matrix values (coherent signaling), while information for the dynamic receiver is transported on the Grassmannian. The static receiver is required to decode and cancel interference, therefore this method is slightly more involved, but it achieves higher DoF compared with the first method.

Using the proposed methods, the exact DoF region is found when $N_1 \leq N_2 \leq M$, $T \geq 2N_1$, where N_1 , N_2 and M are the number of antennas at the dynamic receiver, static receiver and transmitter, respectively, and T is the channel coherence time of the dynamic receiver. For $N_2 < N_1 \leq M$, $T \geq 2N_1$, we partially characterize the DoF region when either the channel is the more capable type [5], or when the message set is degraded [6].

We use the following notation throughout the paper: for a matrix \mathbf{A} , the transpose is denoted with \mathbf{A}^t , the conjugate transpose with \mathbf{A}^\dagger , and the element in row i and column j with $[\mathbf{A}]_{i,j}$. The $k \times k$ identity matrix is denoted with \mathbf{I}_k . The set of $n \times m$ complex matrices is denoted with $\mathcal{C}^{n \times m}$.

The organization of this paper is as follows. In Section II we introduce the system model and preliminary results. Two signaling methods are proposed and studied in Section III and Section IV, respectively.

II. SYSTEM MODEL AND PRELIMINARIES

We consider a broadcast channel with an M -antenna transmitter and two receivers. One receiver has access to channel state information (CSI), and is referred to as the *static receiver*. The other receiver has no CSI, e.g. due to mobility, and is referred to as the *dynamic receiver*. The dynamic receiver has N_1 antennas and the static receiver has N_2 antennas. Denote the channel coefficient matrices from the transmitter to the dynamic and static receivers by $\mathbf{H}_1 \in \mathcal{C}^{N_1 \times M}$ and $\mathbf{H}_2 \in \mathcal{C}^{N_2 \times M}$, respectively. We assume that \mathbf{H}_1 is constant for T symbols (block-fading) and is unknown to both receivers,

Manuscript received November 11, 2011; revised June 6, 2012; accepted June 7, 2012. The material in this paper was presented in part in ISIT 2011 and 2012.

The authors are with the University of Texas at Dallas, Richardson, TX 75080, USA, Email: yang@utdallas.edu, aria@utdallas.edu.

Communicated by Angel Lozano, Editor for Communications.

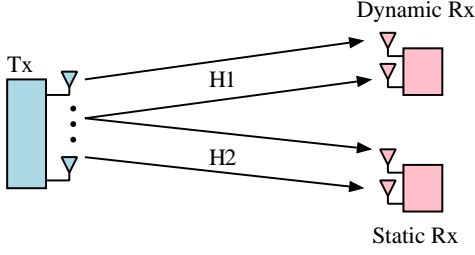


Fig. 1. Channel model.

while \mathbf{H}_2 is known by the static receiver but not known by the dynamic receiver.¹ Neither \mathbf{H}_1 nor \mathbf{H}_2 is known by the transmitter (no CSIT).

Over T time-slots (symbols) the transmitter sends $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_M]^T$ across M antennas, where $\mathbf{x}_i \in \mathbb{C}^{T \times 1}$ is the signal vector sent by the antenna i . The normalized signal at the dynamic and static receivers is respectively

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{H}_1 \mathbf{X} + \frac{1}{\sqrt{\rho}} \mathbf{W}_1, \\ \mathbf{Y}_2 &= \mathbf{H}_2 \mathbf{X} + \frac{1}{\sqrt{\rho}} \mathbf{W}_2, \end{aligned} \quad (1)$$

where $\mathbf{W}_1 \in \mathbb{C}^{N_1 \times T}$ and $\mathbf{W}_2 \in \mathbb{C}^{N_2 \times T}$ are additive noise with i.i.d. entries $\mathcal{CN}(0, 1)$. Each row of $\mathbf{Y}_1 \in \mathbb{C}^{N_1 \times T}$ (or $\mathbf{Y}_2 \in \mathbb{C}^{N_2 \times T}$) corresponds to the received signal at an antenna of the dynamic receiver (or the static receiver) over T time-slots. The transmitter is assumed to have an average power constraint ρ , and therefore, in the normalized channel model given by (1), the average power constraint is:

$$\mathbb{E} \left[\sum_{i=1}^M \text{tr}(\mathbf{x}_i \mathbf{x}_i^\dagger) \right] = T. \quad (2)$$

The channel \mathbf{H}_1 has i.i.d. entries with zero mean and unit variance, but we do *not* assign any specific distribution for \mathbf{H}_1 . This general model includes Rayleigh fading as a special case where the entries of \mathbf{H}_1 are i.i.d. $\mathcal{CN}(0, 1)$. The channel \mathbf{H}_2 is assumed to have full rank; this assumption, e.g., holds with probability 1 if the entries of \mathbf{H}_2 are drawn independently according to a continuous distribution. We focus on the case of $M = \max(N_1, N_2)$ and $T \geq 2N_1$, which is motivated by the fact that having more transmit antennas does not increase the multiplexing gain for either receiver, and the fact that if $T < 2N_1$, some of the antennas of the dynamic receiver can be deactivated without any loss in the degrees of freedom (DoF) [7].

The degrees of freedom at receiver i is defined as:

$$d_i = \lim_{\rho \rightarrow \infty} \frac{R_i(\rho)}{\log \rho},$$

where $R_i(\rho)$ is the rate of the dynamic receiver for $i = 1$ and the static receiver for $i = 2$, respectively.

¹In practice \mathbf{H}_2 for a static receiver may vary across intervals of length much greater than T . However, for the purposes of this paper, once \mathbf{H}_2 is assumed to be known to the static receiver, its time variation (or lack thereof) does not play any role in the subsequent mathematical developments. Therefore in the interest of elegance and for a minimal description of the requirements for the results, we only state that \mathbf{H}_2 is known.

A. Definitions

Definition 1 (Isotropically Distributed Matrix [8]): A random matrix $\mathbf{X} \in \mathbb{C}^{k \times n}$, where $n \geq k$, is called isotropically distributed (i.d.) if its distribution is invariant under unitary transformations, i.e., for any deterministic $n \times n$ unitary matrix Φ ,

$$p(\mathbf{X}) = p(\mathbf{X}\Phi). \quad (3)$$

An example of i.d. matrices is \mathbf{X} with i.i.d. $\mathcal{CN}(0, 1)$ entries.

Remark 1: An interesting property of i.d. matrices is that if \mathbf{X} is i.d. and Φ is a random unitary matrix that is independent of \mathbf{X} , then $\mathbf{X}\Phi$ is independent of Φ [7, Lemma 4]. That is, any rotation to an i.d. matrix is essentially “invisible.”

Definition 2 (Stiefel manifold [9]): The Stiefel manifold $\mathbb{F}(n, k)$, where $n > k$, is the set of all $k \times n$ unitary matrices, i.e.,

$$\mathbb{F}(n, k) = \{\mathbf{Q} \in \mathbb{C}^{k \times n} : \mathbf{Q}\mathbf{Q}^\dagger = \mathbf{I}_k\}.$$

For $k = 1$, the manifold $\mathbb{F}(n, 1)$ is the collection of all n -dimensional vectors with unit norm, i.e., the surface of a unit ball.

Definition 3 (Grassmann manifold [9]): The Grassmann manifold $\mathbb{G}(n, k)$, where $n > k$, is the set of all k -dimensional subspaces of \mathbb{C}^n .

Remark 2: The (complex) dimension of $\mathbb{G}(n, k)$ is

$$\dim(\mathbb{G}(n, k)) = k(n - k), \quad (4)$$

i.e., each point in $\mathbb{G}(n, k)$ has a neighborhood that is equivalent (homeomorphic) to a ball in the Euclidean space of complex dimension $k(n - k)$. The dimensionality of Grassmannian can also be viewed as follows. For any matrix \mathbf{Q} , there exists a $k \times k$ full rank matrix \mathbf{U} so that

$$\mathbf{Q}^* = \mathbf{U}\mathbf{Q} = \begin{bmatrix} 1 & \cdots & 0 & x_{1,k+1} & \cdots & x_{1n} \\ 0 & \cdots & 0 & & & \\ \vdots & & \vdots & & & \vdots \\ 0 & \cdots & 1 & x_{k,k+1} & \cdots & x_{kn} \end{bmatrix}, \quad (5)$$

where \mathbf{Q} and \mathbf{Q}^* span the same row space. Therefore, each point in $\mathbb{G}(n, k)$ is determined by $k(n - k)$ complex parameters x_{ji} , for $1 \leq j \leq k$ and $k + 1 \leq i \leq n$. In other words, a k -dimension subspace in \mathbb{C}^n is uniquely decided by $k(n - k)$ complex variables.

B. Non-coherent Point-to-point Channels

The analysis in this paper uses insights and results from non-coherent communication in point-to-point MIMO channels, which are briefly outlined below.

1) Intuition: Consider a point-to-point $M \times N$ MIMO channel where the receiver does not know the channel \mathbf{H} , namely a non-coherent channel.

At high SNR the additive noise is negligible, so the received signal $\mathbf{Y} \approx \mathbf{H}\mathbf{X}$, where \mathbf{X} is the transmitted signal. Because \mathbf{X} is multiplied by a random and unknown \mathbf{H} , the receiver cannot decode \mathbf{X} . However, communication is still possible because, for any non-singular \mathbf{H} , the received signal \mathbf{Y} spans the same row space as \mathbf{X} . Therefore, the row space of \mathbf{X} can be used to carry information without the need to know \mathbf{H} , i.e., the codebook consists of matrices with different row spaces.

Conveying information via subspaces can be viewed as communication on the Grassmann manifold where each distinct point in the manifold represents a different subspace [7]. In this case, the codewords (information) are represented by subspaces, which differs from the coherent communication that maps each codeword into one point in a Euclidean space [10]. Intuitively, the information of a Grassmannian codeword is carried by $k(n - k)$ variables, as seen in (5).

2) *Optimal Signaling*: The design of an optimal signaling can be viewed as sphere packing over Grassmannians [7]. At high SNR, the optimal signals are isotropically distributed unitary matrices [8], [7]. In addition, the optimal number of transmit antennas depends on the channel coherence time. For a short coherence interval, using fewer antennas may lead to a higher capacity, and vice versa. The optimal number of transmit antennas is

$$K = \min(M, N, \lfloor T/2 \rfloor), \quad (6)$$

where T is the channel coherence time, i.e., the number of symbols that the channel remains constant. Therefore, the optimal signals are $K \times T$ unitary matrices. In other words, K antennas ($K \leq M$) are in use and they transmit equal-energy and mutually orthogonal vectors. These unitary matrices reside in $\mathbb{G}(T, K)$ and each is interpreted as a representation of the subspace it spans. This method achieves the *maximum* DoF $K(T - K)$ over T time-slots. Note that the DoF coincides with the dimensionality of the Grassmannian $\mathbb{G}(T, K)$.

3) *Subspace Decoding*: Unlike coherent communication, in non-coherent signaling the information is embedded in the subspaces instead of the signal values. As long as two matrices span the same subspace, they correspond to the same message. Maximum-likelihood decoding chooses the codeword whose corresponding subspace is the closest one to the subspace spanned by the received signal. For example in [11], the received signals are projected on the subspaces spanned by different codewords, and then the one is chosen with the maximum projection energy. More precisely, for the transmitted signals $\mathbf{X}_i \in \mathcal{C}^{K \times T}$ from a unitary codebook \mathcal{X} , and the received signals $\mathbf{Y} \in \mathcal{C}^{K \times T}$, the ML detector is

$$\hat{\mathbf{X}}_{ML} = \arg \max_{\mathbf{X}_i \in \mathcal{X}} \text{tr}\{\mathbf{Y}\mathbf{X}_i^\dagger \mathbf{X}_i \mathbf{Y}^\dagger\}. \quad (7)$$

C. A Baseline Scheme: Orthogonal Transmission

For the purposes of establishing a baseline for comparison, we begin by considering a time-sharing (orthogonal transmission) that acquires CSIR via training in each interval and uses Gaussian signaling. This baseline method has been chosen to highlight the differences of the heterogeneous MIMO broadcast channel of this paper with two other known scenarios: It is known that for a broadcast channel with no CSIT and perfect CSIR, orthogonal transmission achieves the optimal DoF region [2]. Also, a training-based method with Gaussian signaling is sufficient to achieve DoF optimality [7] for the point-to-point noncoherent MIMO channel².

²Grassmannian signaling is superior, but the same slope of the rate vs. SNR curve is obtained with training and Gaussian signaling in the point-to-point MIMO channel.

In orthogonal transmission, the transmitter communicates with the two receivers in a time-sharing manner. When transmitting to the dynamic receiver, it is optimal if the transmitter activates only K out of M antennas: it sends pilots from the K antennas sequentially over the first K time-slots; the dynamic receiver estimates the channel by using, e.g., minimum-mean-square-error (MMSE) estimation. Then, the transmitter sends data during the remaining $(T - K)$ time-slots, and the dynamic receiver decodes the data by using the estimated channel coefficients [7], [12]. Using this strategy, the maximum rate achieved by the dynamic receiver is:

$$K(1 - \frac{K}{T}) \log \rho + O(1). \quad (8)$$

The operating point in the achievable DoF region where the transmitter communicates exclusively with the dynamic receiver is denoted with \mathcal{D}_1 .

$$\mathcal{D}_1 = (K(1 - \frac{K}{T}), 0). \quad (9)$$

For the static receiver the channel is assumed to be known at the receiver, therefore data is transmitted to it coherently. The maximum rate achieved by the static receiver is [13]

$$\min(M, N_2) \log \rho + O(1). \quad (10)$$

The operating point in the DoF region where the transmitter communicates only with the static receiver is denoted with \mathcal{D}_2 .

$$\mathcal{D}_2 = (0, \min(M, N_2)). \quad (11)$$

Time-sharing between the two points of \mathcal{D}_1 and \mathcal{D}_2 yields the achievable DoF region

$$\left(tK(1 - \frac{K}{T}), (1 - t) \min(M, N_2) \right), \quad (12)$$

where t is a time-sharing variable.

III. GRASSMANNIAN SUPERPOSITION FOR BROADCAST CHANNEL

In this section, we propose a signaling method that attains DoF region superior to orthogonal transmission, and allows each receiver to decode its message while being oblivious of the other receiver's message.

A. A Toy Example

Consider $M = N_2 = 2$, $N_1 = 1$ and $T = 2$. From Section II-C, orthogonal transmission attains 1/2 DoF per time-slot for the dynamic receiver and 2 DoF per time-slot for the static receiver. By time-sharing between the two receivers, the following DoF region is achieved

$$(\frac{t}{2}, 2 - 2t), \quad (13)$$

where $t \in [0, 1]$ is a time-sharing parameter.

We now consider the transmitter sends a product of signal vectors over 2 time-slots

$$\mathbf{X} = \mathbf{x}_2 \mathbf{x}_1^t \in \mathcal{C}^{2 \times 2}, \quad (14)$$

where $\mathbf{x}_1 = [x_1^{(1)} \ x_2^{(1)}]^t$ and $\mathbf{x}_2 = [x_1^{(2)} \ x_2^{(2)}]^t$ are the signals for the dynamic receiver and the static receiver, respectively. The vectors \mathbf{x}_1 and \mathbf{x}_2 have unit-norm and from codebooks that lie on $\mathbb{G}(2, 1)$.

The signal at the dynamic receiver is

$$\begin{aligned} \mathbf{y}_1 &= [h_1^{(1)} \ h_2^{(1)}] \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \end{bmatrix} [x_1^{(1)} \ x_2^{(1)}] + \frac{1}{\sqrt{\rho}} [w_1^{(1)} \ w_2^{(1)}] \\ &= \tilde{h}^{(1)} [x_1^{(1)} \ x_2^{(1)}] + \frac{1}{\sqrt{\rho}} [w_1^{(1)} \ w_2^{(1)}], \end{aligned} \quad (15)$$

where $[h_1^{(1)}, h_2^{(1)}]$ is the isotropically distributed channel vector, and $\tilde{h}^{(1)}$ is the equivalent channel coefficient seen by the dynamic receiver.

The subspace spanned by \mathbf{x}_1^t is the same as $\tilde{h}^{(1)} \mathbf{x}_1^t$, so at high SNR the dynamic receiver is able to determine the direction specified by \mathbf{x}_1^t . From Section II-B, the dynamic receiver attains 1/2 DoF per time-slot, which is optimal even in the absence of the static receiver.

Consider the signal of the static receiver at time-slot 1:

$$\mathbf{y}_2 = \mathbf{H}_2 \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \end{bmatrix} x_1^{(1)} + \frac{1}{\sqrt{\rho}} \begin{bmatrix} w_1^{(2)} \\ w_2^{(2)} \end{bmatrix}. \quad (16)$$

Because the static receiver knows \mathbf{H}_2 , it can invert the channel³ as long as \mathbf{H}_2 is non-singular:

$$(\mathbf{H}_2^{-1} \mathbf{y}_2)^t = x_1^{(1)} [x_1^{(2)} \ x_2^{(2)}] + [w_1^{(2)} \ w_2^{(2)}] \mathbf{H}_2^{-t}. \quad (17)$$

The equivalent (unknown) channel seen by the static receiver is $x_1^{(1)}$, i.e., part of the dynamic receiver's signal. Using Grassmannian signaling via the subspace of \mathbf{x}_2 , the DoF achieved is again 1/2 per time-slot.

Time-sharing between the proposed scheme and \mathcal{D}_2 (transmitting only to the static receiver) yields the achievable DoF region

$$\left(\frac{1}{2}t, \ 2 - \frac{3}{2}t \right). \quad (18)$$

The above region is strictly larger than that of orthogonal transmission, as shown in Figure 2. The static receiver achieves 1/2 DoF “for free” in the sense that this DoF was extracted for the static receiver without reducing the dynamic receiver's DoF.

B. Grassmannian Superposition Signaling

Based on the previous example, we design a general signaling method (the Grassmannian superposition) with two properties: (1) information is carried by subspaces and (2) two signal matrices are superimposed multiplicatively so that their row (or column) space is unaffected by multiplying the other receiver's signal matrix. Two separate cases are considered based on whether the number of static receiver antennas is larger than the number of dynamic receiver antennas.

³The noise enhancement induced by channel inversion will not affect the DoF of the static receiver.

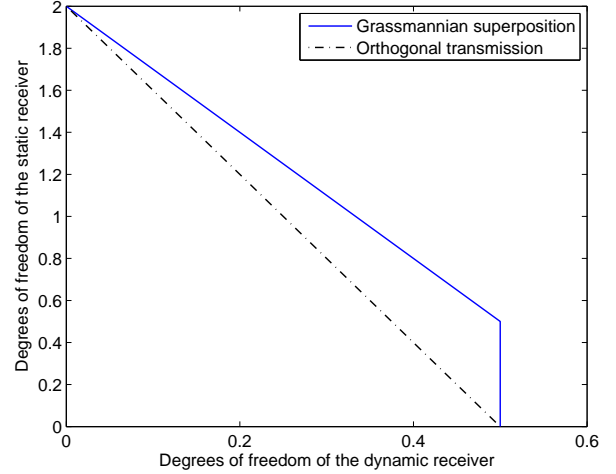


Fig. 2. DoF region of the toy example 1.

1) $N_1 < N_2$: The transmitter sends $\mathbf{X} \in \mathcal{C}^{N_2 \times T}$ across $M = N_2$ antennas over an interval of length T :

$$\mathbf{X} = \sqrt{\frac{T}{N_1}} \mathbf{X}_2 \mathbf{X}_1, \quad (19)$$

where $\mathbf{X}_1 \in \mathcal{C}^{N_1 \times T}$ and $\mathbf{X}_2 \in \mathcal{C}^{N_2 \times N_1}$ are the signals for the dynamic receiver and the static receiver, respectively. Here, $\sqrt{T/N_1}$ is a normalizing factor to satisfy the power constraint (2). Information for both receivers are sent over the Grassmannian, namely \mathbf{X}_1 is from a codebook $\mathcal{X}_1 \subset \mathbb{G}(T, N_1)$ and \mathbf{X}_2 is from a codebook $\mathcal{X}_2 \subset \mathbb{G}(N_2, N_1)$. The codebook \mathcal{X}_1 and \mathcal{X}_2 are chosen to be isotropically distributed unitary matrices (see Section III-C for more details).

A sketch of the argument for the DoF achieved by the Grassmannian superposition is as follows. The noise is negligible at high SNR, so the signal at the dynamic receiver is approximately

$$\mathbf{Y}_1 \approx \sqrt{\frac{T}{N_1}} \mathbf{H}_1 \mathbf{X}_2 \mathbf{X}_1 \in \mathcal{C}^{N_1 \times T}. \quad (20)$$

The row space of \mathbf{X}_1 can be determined based on \mathbf{Y}_1 , and then $(T - N_1)N_1$ independent variables (DoF) that specify the row space are recovered, i.e., the transmitted point \mathbf{X}_1 in $\mathcal{X}_1 \in \mathbb{G}(T, N_1)$ is found.

For the static receiver, since \mathbf{H}_2 is known by the receiver, it inverts the channel (given that \mathbf{H}_2 is non-singular)

$$\mathbf{H}_2^{-1} \mathbf{Y}_2 \approx \sqrt{\frac{T}{N_1}} \mathbf{X}_2 \mathbf{X}_1 \in \mathcal{C}^{N_2 \times T}, \quad (21)$$

which has approximately the same column space as \mathbf{X}_2 . The transmitted point \mathbf{X}_2 in $\mathcal{X}_2 \in \mathbb{G}(N_2, N_1)$ will be recovered from the column space of $\mathbf{H}_2^{-1} \mathbf{Y}_2$, producing $(N_2 - N_1)N_1$ DoF.

Therefore, the proposed scheme attains the DoF pair

$$\mathcal{D}_3 = \left(N_1 \left(1 - \frac{N_1}{T} \right), \frac{N_1}{T} (N_2 - N_1) \right). \quad (22)$$

The result is more formally stated as follows:

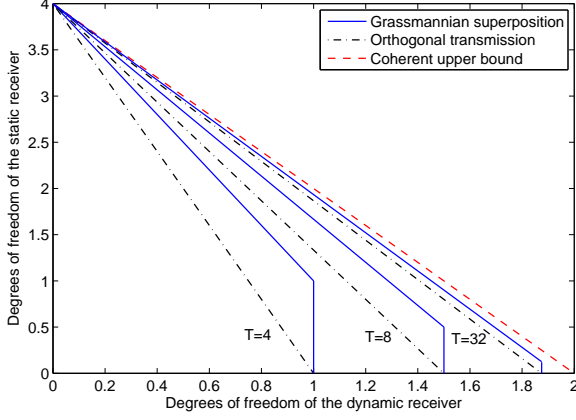


Fig. 3. DoF region (Theorem 1): $N_1 = 2$, $N_2 = 4$.

Theorem 1 ($N_1 < N_2$): Consider a broadcast channel with an M -antenna transmitter, a dynamic receiver and a static receiver with N_1 and N_2 antennas, respectively, with coherence time T for the dynamic channel. The Grassmannian superposition achieves the rate pair

$$\begin{cases} R_1 = N_1 \left(1 - \frac{N_1}{T}\right) \log \rho + O(1) \\ R_2 = \frac{N_1}{T} (N_2 - N_1) \log \rho + O(1) \end{cases}.$$

The corresponding DoF pair is denoted

$$\mathcal{D}_3 = \left(N_1 \left(1 - \frac{N_1}{T}\right), \frac{N_1}{T} (N_2 - N_1) \right).$$

If we denote the DoF for the single-user operating points for the dynamic and static user with \mathcal{D}_1 , \mathcal{D}_2 respectively, the achievable DoF region consists of the convex hull of \mathcal{D}_1 , \mathcal{D}_2 and \mathcal{D}_3 .

Proof: See Appendix I. ■

From Theorem 1 the static receiver attains a “free” rate of

$$\Delta R_1 = \frac{N_1}{T} (N_2 - N_1) \log \rho + O(1). \quad (23)$$

We plot the achievable DoF region of Theorem 1 in Figure 3. For small T , the DoF gain achieved by the proposed method is significant, while as T increases, both methods approach the coherent upper bound [2] where both of the receivers have CSIR. For $T \rightarrow \infty$, the rate gain $\Delta R_1 = O(1)$, and no DoF gain is obtained. In this case, the achievable DoF region in Theorem 1 coincides with that attained by orthogonal transmission as well as the coherent outer bound [2]. This is not surprising, since if the channel remains constant ($T \rightarrow \infty$), the resource used for obtaining CSIR is negligible. Finally, the rate gain ΔR_1 is an increasing function of $(N_2 - N_1)$, i.e., the extra antennas available for the static receiver.

Now, we design the dimension of \mathbf{X}_1 and \mathbf{X}_2 in (19) to maximize the achievable DoF region. To find the optimal dimensions, we allow the signaling to use a flexible number of antennas and time slots, up to the maximum available. Let $\mathbf{X}_1 \in \mathcal{C}^{N_1 \times \hat{T}}$ and $\mathbf{X}_2 \in \mathcal{C}^{N_2 \times \hat{N}_1}$, where $\hat{T} \leq T$, $\hat{N}_1 \leq N_1$ and $\hat{N}_2 \leq N_2$. Theorem 1 does not immediately reveal the optimal values of \hat{N}_1 , \hat{N}_2 , and \hat{T} , because the rates are not monotonic

in the mentioned parameters. The following corollary presents the optimal value of \hat{N}_1 , \hat{N}_2 and \hat{T} .

Corollary 1: For the Grassmannian superposition under $N_1 < N_2$, the signal dimension $\hat{T} = T$, $\hat{N}_1 = N_1$ and $\hat{N}_2 = N_2$ optimizes the achievable DoF region.

Proof: See Appendix II. ■

Thus, in the special case of $N_1 < N_2$, it is optimal to use all time slots and all antennas.

2) $N_1 \geq N_2$: In this case, we shall see that sometimes the Grassmannian superposition may still outperform orthogonal transmission, but also under certain conditions (e.g. very large T or $N_1 \gg N_2$) the Grassmannian superposition as described in this section may not improve the DoF compared with orthogonal transmission.

When $N_1 \geq N_2$, if the Grassmannian signaling to the dynamic receiver uses all the N_1 dimensions, there will remain no room for communication with the static receiver. To allow the static user to also use the channel, the dynamic user must “back off” from using all the rate available to it, in other words, the dimensionality of the signaling for the dynamic receiver must be reduced. The largest value of \hat{N}_1 that makes $\hat{N}_1 < N_2$ and thus allows nontrivial Grassmannian superposition is $\hat{N}_1 = N_2 - 1$. Once we are in this regime, the results of the subsection III-B.1 can be used. Specifically, Corollary 1 indicates that de-activating any further dynamic user antennas will not improve the DoF region. Thus, given N_2 , and assuming we wish to have a non-trivial Grassmannian signaling for both users, using $\hat{N}_1 = N_2 - 1$ dimensions for signaling to the dynamic receiver maximizes the DoF region. The transmit signal is then

$$\mathbf{X} = \sqrt{\frac{T}{N_1}} \mathbf{X}_2 \mathbf{X}_1, \quad (24)$$

where $\mathbf{X}_1 \in \mathcal{C}^{(N_2-1) \times T}$ and $\mathbf{X}_2 \in \mathcal{C}^{N_2 \times (N_2-1)}$. The corresponding achievable DoF pair is

$$\mathcal{D}_4 = \left((N_2 - 1) \left(1 - \frac{N_2 - 1}{T}\right), (N_2 - 1)/T \right), \quad (25)$$

which leads to the following result.

Corollary 2 ($N_1 \geq N_2$): Consider an M -antenna transmitter broadcasting to a dynamic receiver and a static receiver with N_1 and N_2 antennas, respectively, with coherence time T for the dynamic channel. Then the Grassmannian superposition achieves the rate pair

$$\begin{cases} R_1 = (N_2 - 1) \left(1 - \frac{N_2 - 1}{T}\right) \log \rho + O(1) \\ R_2 = \frac{N_2 - 1}{T} \log \rho + O(1) \end{cases}.$$

Denote the corresponding DoF pair with \mathcal{D}_4 . Together with the two single-user operating points \mathcal{D}_1 and \mathcal{D}_2 obtained earlier, the achievable DoF region consists of the convex hull of \mathcal{D}_1 , \mathcal{D}_2 and \mathcal{D}_4 .

Proof: The proof follows directly by replacing N_1 with $(N_2 - 1)$ in Theorem 1. ■

In Corollary 2, the DoF for the static receiver has not been achieved for “free” but at the expense of reducing the DoF for the dynamic receiver. The transmitter uses only $N_2 - 1$ dimensions for the dynamic receiver, which allows an extra DoF $(N_2 - 1)/T$ to be attained at the static receiver. If $N_1 - N_2$

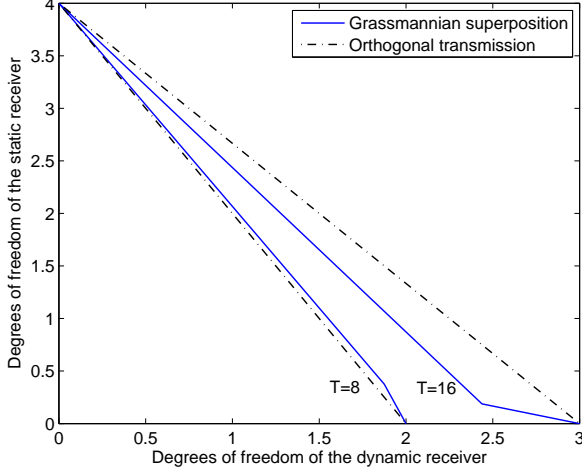


Fig. 4. DoF region (Corollary 2): $N_1 = N_2 = 4$.

and T are small, then the DoF gain of the static receiver outweighs the DoF loss for the dynamic, so that the overall achievable DoF region will be superior to that of orthogonal transmission. In contrast, if $N_1 \gg N_2$ or T is large, the DoF loss from the dynamic receiver may not be compensated by the DoF gain from the static receiver, as illustrated by Figure 4. Therefore in the latter case orthogonal transmission may do better. The following corollary specifies the condition under which Grassmannian superposition improves DoF region compared with orthogonal transmission.

Corollary 3: For $N_1 \geq N_2$, the Grassmannian superposition improves DoF region with respect to orthogonal transmission if and only if

$$\frac{N_2 - (N_2 - 1)/T}{(N_2 - 1)(1 - (N_2 - 1)/T)} < \frac{N_2}{N_1(1 - N_1/T)}. \quad (26)$$

Proof: The necessary and sufficient condition for ensuring the improvement of the achievable DoF region is that the slope between \mathcal{D}_2 and \mathcal{D}_4 is larger than the slope between \mathcal{D}_1 and \mathcal{D}_2 , which is equivalent to the inequality in the corollary. ■

C. Design of \mathcal{X}_1 and \mathcal{X}_2

The representation of a point in the Grassmannian is not unique [14] (also see Remark 2), and therefore the codebooks $\mathcal{X}_1 \subset \mathbb{G}(T, N_1)$ and $\mathcal{X}_2 \subset \mathbb{G}(N_2, N_1)$ are not unique.

First, \mathcal{X}_2 is chosen to be a unitary codebook. When \mathbf{X}_2 is unitary, for i.i.d. Rayleigh fading \mathbf{H}_1 , the equivalent dynamic channel $\tilde{\mathbf{H}}_1 = \mathbf{H}_1 \mathbf{X}_2$ still has i.i.d. Rayleigh fading coefficients [8]. Therefore, the static receiver is *transparent* to the dynamic receiver, which allows us to decouple and simplify the design of the two codebooks and their decoders.

Once \mathcal{X}_2 is chosen to be a set of unitary matrices, communication between dynamic receiver and the transmitter is equivalent to a non-coherent point-to-point MIMO channel. Hence, to maximize the rate of the dynamic receiver at high SNR, \mathcal{X}_1 must also be a collection of isotropically distributed unitary matrices (see Section II).

Remark 3: With unitary codebooks \mathcal{X}_1 and \mathcal{X}_2 , information for both receivers is conveyed purely by the subspace to which the codeword belongs. Consider $\mathbf{X} \in \mathbb{C}^{k \times n}$, $n \geq k$, which is uniquely represented by Ω (the row space of \mathbf{X}) and a $k \times k$ coefficient matrix \mathbf{C} according to a certain basis of Ω . The codewords $\mathbf{X}_1, \mathbf{X}_2$ can be represented as

$$\begin{aligned} \mathbf{X}_1 &\rightarrow (\Omega_1, \mathbf{C}_1), \\ \mathbf{X}_2 &\rightarrow (\Omega_2, \mathbf{C}_2). \end{aligned} \quad (27)$$

In a manner similar to [7], one can verify

$$I(\mathbf{X}_1; \mathbf{Y}_1) = I(\Omega_1; \mathbf{Y}_1) + \underbrace{I(\mathbf{C}_1; \mathbf{Y}_1 | \Omega_1)}_{=0}, \quad (28)$$

and

$$\begin{aligned} I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{H}_2) &= I(\Omega_2; \mathbf{Y}_2 | \mathbf{H}_2) \\ &\quad + \underbrace{I(\mathbf{C}_2; \mathbf{Y}_2 | \Omega_2, \mathbf{H}_2)}_{=0}. \end{aligned} \quad (29)$$

D. Multiplicative vs. Additive Superposition

In this section, we compare product superposition with additive superposition. Under additive superposition, the transmit signal has a general expression

$$\mathbf{X} = \sqrt{c_1 \rho} \mathbf{V}_1 \mathbf{X}_1 + \sqrt{c_2 \rho} \mathbf{V}_2 \mathbf{X}_2, \quad (30)$$

where \mathbf{V}_1 and \mathbf{V}_2 are the precoding matrices, and c_1 and c_2 represent the power allocation. In this case, the signal at the dynamic receiver is

$$\mathbf{Y}_1 = \sqrt{c_1 \rho} \mathbf{H}_1 \mathbf{V}_1 \mathbf{X}_1 + \sqrt{c_2 \rho} \mathbf{H}_1 \mathbf{V}_2 \mathbf{X}_2 + \mathbf{W}_1. \quad (31)$$

Since \mathbf{H}_1 is unknown, the second interference term cannot be completely eliminated in general, which leads to a bounded signal-to-interference-plus-noise ratio (SINR), resulting in zero DoF for the dynamic receiver.

For the multiplicative superposition, the signal at the dynamic receiver is

$$\begin{aligned} \mathbf{Y}_1 &= \sqrt{c \rho} \mathbf{H}_1 \mathbf{X}_2 \mathbf{X}_1 + \mathbf{W}_1 \\ &= \sqrt{c \rho} \tilde{\mathbf{H}}_1 \mathbf{X}_1 + \mathbf{W}_1, \end{aligned} \quad (32)$$

where c is a power normalizing constant. For any unitary \mathbf{X}_2 , $\mathbf{X}_2 \mathbf{X}_1$ and \mathbf{X}_1 span the same row space. This invariant property of Grassmannian enables us to convey information to the static receiver via \mathbf{X}_2 without reducing the degrees of freedom of the dynamic receiver. Intuitively, the dynamic receiver does not have CSIR and is “insensitive” to rotation, i.e., the distribution of \mathbf{Y}_1 does not depend on \mathbf{X}_2 .

For the static receiver, the received signal is

$$\mathbf{Y}_2 = \sqrt{c \rho} \mathbf{H}_2 \mathbf{X}_2 \mathbf{X}_1 + \mathbf{W}_2. \quad (33)$$

Because \mathbf{H}_2 is known, the channel rotation \mathbf{X}_2 is detectable, i.e., the distribution of \mathbf{Y}_2 depends on \mathbf{X}_2 . Therefore \mathbf{X}_2 can be used to convey information for the static receiver.

IV. GRASSMANNIAN-EUCLIDEAN SUPERPOSITION FOR THE BROADCAST CHANNEL

We now propose a new transmission scheme based on successive interference cancellation, where the static receiver decodes and removes the signal for the dynamic receiver before decoding its own signal. This scheme improves the DoF region compared to the non-interfering Grassmannian signaling of the previous section.

A. A Toy Example

Consider $M = N_1 = N_2 = 1$ and $T = 2$. Our approach is that over 2 time-slots, the transmitter sends

$$\mathbf{x} = x_2 \mathbf{x}_1^t \in \mathbb{C}^{1 \times 2}, \quad (34)$$

where $\mathbf{x}_1 = [x_1^{(1)} \ x_2^{(1)}]^t$ is the signal for the dynamic receiver and x_2 is the signal for the static receiver. Here, \mathbf{x}_1 has unit-norm and is from a codebook \mathcal{X}_1 that is a subset of $\mathbb{G}(2, 1)$, and x_2 can obey any distribution that satisfies the average power constraint.

The signal at the dynamic receiver is

$$\mathbf{y}_1 = h_1 x_2 [x_1^{(1)} \ x_2^{(1)}] + \frac{1}{\sqrt{\rho}} [w_1^{(1)} \ w_2^{(1)}] \quad (35)$$

$$= \tilde{h}_1 [x_1^{(1)} \ x_2^{(1)}] + \frac{1}{\sqrt{\rho}} [w_1^{(1)} \ w_2^{(1)}], \quad (36)$$

where h_1 is the channel coefficient of the dynamic receiver, and $\tilde{h}_1 \triangleq h_1 x_2$ is the equivalent channel coefficient. The dynamic receiver can determine the row space spanned by \mathbf{x}_1 even though \tilde{h}_1 is unknown, in a manner similar to Section III-A. The total DoF conveyed by \mathbf{x}_1 is 1 (thus $\frac{1}{2}$ per time-slot); this is the optimal DoF under the same number of antennas and coherence time.

For the static receiver, the received signal is:

$$\mathbf{y}_2 = h_2 x_2 [x_1^{(1)} \ x_2^{(1)}] + \frac{1}{\sqrt{\rho}} [w_1^{(2)} \ w_2^{(2)}] \quad (37)$$

$$= \tilde{h}_2 [x_1^{(1)} \ x_2^{(1)}] + \frac{1}{\sqrt{\rho}} [w_1^{(2)} \ w_2^{(2)}], \quad (38)$$

where h_2 is the channel coefficient of the static receiver, and $\tilde{h}_2 \triangleq h_2 x_2$. Intuitively, since (36) and (38) are equivalent, if the dynamic receiver decodes the subspace of \mathbf{x}_1 , so does the static receiver. Then, the exact signal vector \mathbf{x}_1 is known to the static receiver (recall that each subspace is uniquely represented by a signal matrix). The static receiver removes the interference signal \mathbf{x}_1

$$\mathbf{y}_2 \mathbf{x}_1^\dagger = h_2 x_2 + \frac{1}{\sqrt{\rho}} \tilde{w}_2, \quad (39)$$

where \tilde{w}_2 is the equivalent noise. Finally, the static receiver knows h_2 , so it decodes x_2 and attains $1/2$ DoF per time-slot.

Therefore, the proposed scheme attains the maximum DoF for the dynamic receiver, meanwhile achieving $1/2$ DoF for the static receiver. With time sharing between this scheme and \mathcal{D}_2 , the achievable DoF pair is

$$(d_1, d_2) = \left(\frac{t}{2}, 1 - \frac{t}{2} \right). \quad (40)$$

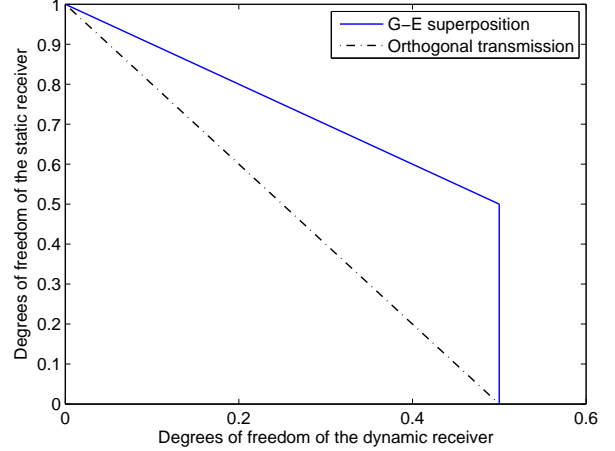


Fig. 5. DoF region of the toy example 2.

Figure 5 shows that this region is uniformly larger than that of orthogonal transmission.

Remark 4: There are two key differences between the method proposed here and the Grassmannian superposition proposed in Section III. First, the information for the static receiver is carried by the *value* of x_2 instead of its direction (subspace), i.e., the signal for the static receiver is carried in the Euclidean space. Second, the static receiver must decode and remove the interference signal for the dynamic receiver before decoding its own signal, which is unlike the non-interfering method of the previous section.

B. Grassmannian-Euclidean Superposition Signaling

We denote the aforementioned method as *Grassmannian-Euclidean superposition*, whose generalization is the subject of this subsection. Two separate cases are considered based on whether the number of static receiver antennas is less than, or no less than, the number of dynamic receiver antennas.

1) $N_1 \leq N_2$: The transmitter sends $\mathbf{X} \in \mathbb{C}^{N_2 \times T}$

$$\mathbf{X} = \sqrt{\frac{T}{N_1 N_2}} \mathbf{X}_2 \mathbf{X}_1, \quad (41)$$

where $\mathbf{X}_1 \in \mathbb{C}^{N_1 \times T}$ and $\mathbf{X}_2 \in \mathbb{C}^{N_2 \times N_1}$ are signals for the dynamic receiver and the static receiver, respectively. The signal \mathbf{X}_1 is from a Grassmannian codebook $\mathcal{X}_1 \subset \mathbb{G}(T, N_1)$, while \mathbf{X}_2 is from a conventional Gaussian codebook \mathcal{X}_2 . The constant $\sqrt{T/N_1 N_2}$ is a power normalizing factor.

We now give a sketch of the argument of the DoF attained by the superposition signaling (41). For the dynamic receiver, $\mathbf{Y}_1 \approx \mathbf{H}_1 \mathbf{X}_2 \mathbf{X}_1$ at high SNR. When $N_1 \leq N_2$, the equivalent channel $\mathbf{H}_1 \mathbf{X}_2 \in \mathbb{C}^{N_1 \times N_1}$ has full rank and does not change the row space of \mathbf{X}_1 . Recovering the row space of \mathbf{X}_1 produces $(T - N_1)N_1$ DoF, which is similar to Section III.

For the static receiver, the signal at high SNR is

$$\mathbf{Y}_2 \approx \sqrt{\frac{T}{N_1 N_2}} \mathbf{H}_2 \mathbf{X}_2 \mathbf{X}_1 = \sqrt{\frac{T}{N_1 N_2}} \tilde{\mathbf{H}}_2 \mathbf{X}_1. \quad (42)$$

For $N_1 \leq N_2$, $\tilde{\mathbf{H}}_2 = \mathbf{H}_2 \mathbf{X}_2 \in \mathbb{C}^{N_2 \times N_1}$ has full column rank and does not change the row space of \mathbf{X}_1 , and therefore,

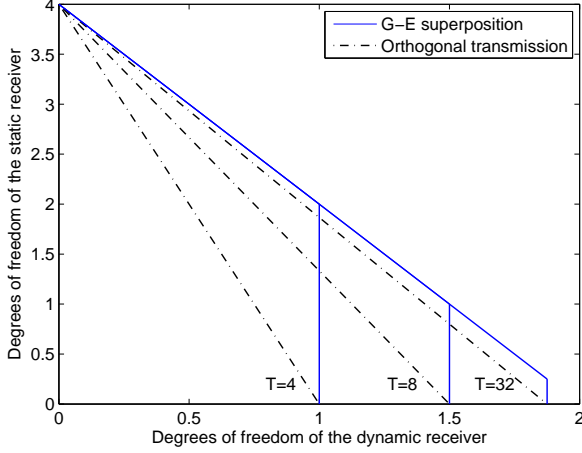


Fig. 6. DoF region (Theorem 2): $N_1 = 2$, $N_2 = 4$.

the signal intended for the dynamic receiver can be decoded by the static receiver. From the subspace spanned by \mathbf{X}_1 , the codeword $\mathbf{X}_1 \in \mathcal{X}_1$ is identified. Then, \mathbf{X}_1 is peeled off from the static signal:

$$\mathbf{Y}_2 \mathbf{X}_1^\dagger \approx \sqrt{\frac{T}{N_1 N_2}} \mathbf{H}_2 \mathbf{X}_2 \in \mathcal{C}^{N_2 \times N_1}. \quad (43)$$

Because \mathbf{H}_2 is known by the static receiver, Eq. (43) is a point-to-point MIMO channel. Therefore, $N_2 N_1$ DoF can be communicated via \mathbf{X}_2 to the static receiver (over T time-slots) [13].

Altogether, the Grassmannian-Euclidean superposition attains the DoF pair \mathcal{D}_5

$$\mathcal{D}_5 = \left(N_1(1 - N_1/T), N_2 N_1/T \right). \quad (44)$$

More precisely, we have the following theorem.

Theorem 2 ($N_1 \leq N_2$): Consider a broadcast channel with an M -antenna transmitter, a dynamic receiver and a static receiver with N_1 and N_2 antennas, respectively, with coherence time T for the dynamic channel. The Grassmannian-Euclidean superposition achieves the rate pair

$$\begin{cases} R_1 = N_1(1 - \frac{N_1}{T}) \log \rho + O(1) \\ R_2 = \frac{N_1 N_2}{T} \log \rho + O(1) \end{cases}.$$

Denote the corresponding DoF pair by \mathcal{D}_5 . Together with the two single-user operating points \mathcal{D}_1 , \mathcal{D}_2 obtained earlier, the achievable DoF region consists of the convex hull of \mathcal{D}_1 , \mathcal{D}_2 and \mathcal{D}_5 .

Proof: See Appendix III. ■

With the Grassmannian-Euclidean superposition, the static receiver attains the following gain compared with orthogonal transmission:

$$\Delta R_2 = \frac{N_1 N_2}{T} \log \rho + O(1). \quad (45)$$

From Figure 6, for relatively small T or large N_2 , the DoF gain is significant. For example, at $T = 2N_1$, the minimum coherence interval considered in this paper, the proposed method achieves a DoF $N_2/2$ for the static receiver while

attaining the maximum DoF $N_1/2$ for the dynamic receiver. As T increases the gain over orthogonal transmission decreases. In the limit $T \rightarrow \infty$, we have $\Delta R_2 = O(1)$, and the DoF gain of Grassmannian-Euclidean superposition goes away. The Grassmannian-Euclidean superposition also provides DoF gain over the non-interfering Grassmannian superposition⁴

$$\Delta R = \frac{N_1^2}{T} \log \rho + O(1). \quad (46)$$

The optimal design of the dimensions of \mathbf{X}_1 and \mathbf{X}_2 is trivial, because the DoF region in Theorem 2 is indeed optimal (see Section IV-D).

2) $N_1 > N_2$: When the static receiver has fewer antennas than the dynamic receiver, it may not be able to decode the dynamic signal. Here, we cannot directly apply the signaling structure given by (41). A straightforward way is to activate only N_2 antennas at the transmitter and use only N_2 dimensions for the dynamic receiver, that is

$$\mathbf{X} = \sqrt{\frac{T}{N_2^2}} \mathbf{X}_2 \mathbf{X}_1 \in \mathcal{C}^{N_2 \times T}, \quad (47)$$

where $\mathbf{X}_1 \in \mathcal{C}^{N_2 \times T}$ and $\mathbf{X}_2 \in \mathcal{C}^{N_2 \times N_2}$, and $\sqrt{T/N_2^2}$ is a power normalizing factor.

Following the same argument as the case of $N_1 \leq N_2$, the Grassmannian-Euclidean superposition achieves the DoF pair

$$\mathcal{D}_6 = \left(N_2(1 - \frac{N_2}{T}), \frac{N_2^2}{T} \right). \quad (48)$$

Corollary 4 ($N_1 > N_2$): Consider a broadcast channel with an M -antenna transmitter, a dynamic receiver and a static receiver with N_1 and N_2 antennas, respectively, with coherence time T for the dynamic channel. The Grassmannian-Euclidean superposition achieves the rate pair

$$\begin{cases} R_1 = N_2(1 - \frac{N_2}{T}) \log \rho + O(1) \\ R_2 = \frac{N_2^2}{T} \log \rho + O(1) \end{cases}$$

Denote the corresponding DoF pair with \mathcal{D}_6 . Together with the two single-user operating points \mathcal{D}_1 and \mathcal{D}_2 obtained earlier, the achievable DoF region consists of the convex hull of \mathcal{D}_1 , \mathcal{D}_2 and \mathcal{D}_6 .

Proof: The proof directly follows from Theorem 2. ■

In Corollary 4, the static rate receiver is obtained at the expense of a reduction in the dynamic rate. The transmitter uses only N_2 out of N_1 dimensions available for the dynamic receiver, which allows extra DoF N_2^2/T for the static receiver. A necessary and sufficient condition for Grassmannian-Euclidean superposition to improve the DoF region is as follows.

Corollary 5: For the Grassmannian-Euclidean superposition, the signal dimension $\hat{T} = T$, $\hat{N}_1 = N_1$ and $\hat{N}_2 = N_2$ optimizes the rate region at high SNR. Moreover, it achieves superior DoF region compared with orthogonal transmission if and only if

$$N_2 > (1 - \frac{N_1}{T}) N_1 \quad (49)$$

⁴ Although Grassmannian-Euclidean superposition achieves larger DoF than the non-interfering Grassmannian superposition, it may not achieve larger rate at low or moderate SNR due to the decodable restriction on the rate (interference).

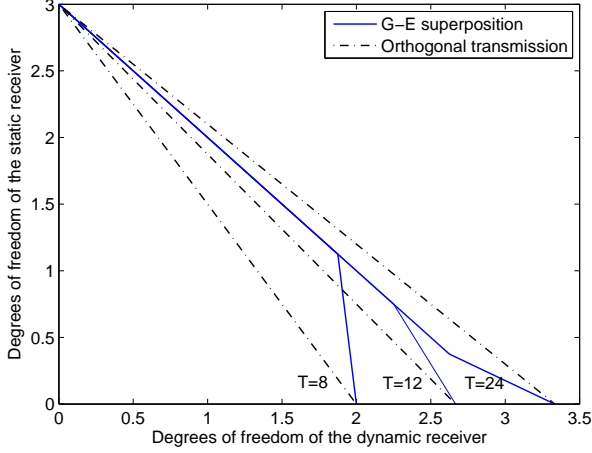


Fig. 7. DoF region (Corollary 4): $N_1 = 4$, $N_2 = 3$.

Proof: First, using the maximum number of static antennas ($\hat{N}_2 = N_2$) is optimal, because both R_1 and R_2 in Corollary 4 are increasing functions of N_2 (note that $N_2 \leq T/2$).

Second, we find the optimal \hat{T} . Maximizing the achievable DoF region is equivalent to maximizing the slope of the line between \mathcal{D}_2 and \mathcal{D}_4 , i.e.,

$$(0, N_2) \quad \text{and} \quad \left(N_2\left(1 - \frac{N_2}{\hat{T}}\right), \frac{N_2^2}{\hat{T}}\right), \quad (50)$$

which has a constant slope -1 and is independent of \hat{T} . Therefore, any choice of \hat{T} , as long as $\hat{T} \geq 2N_2$, achieves a boundary point of the DoF region of the Grassmannian-Euclidean superposition.

Finally, for the Grassmannian-Euclidean superposition to be superior to orthogonal transmission in term of DoF, the slope of the line between \mathcal{D}_2 and \mathcal{D}_6 must be larger than the slope between \mathcal{D}_1 and \mathcal{D}_2 , namely

$$\frac{N_2}{(1 - N_1/T)N_1} > 1. \quad (51)$$

This completes the proof. \blacksquare

Corollary 5 can be interpreted as follows: the Grassmannian-Euclidean superposition achieves superior DoF if and only if the maximum DoF of the static receiver is larger than that of the dynamic receiver.

C. Design of \mathcal{X}_1 and \mathcal{X}_2

We heuristically argue that it is reasonable to choose \mathcal{X}_1 to be isotropically distributed unitary matrices and \mathcal{X}_2 to be i.i.d. complex Gaussian codebook.

Recall that the Grassmannian-Euclidean superposition is to allow the static receiver to decode the signal for the dynamic receiver and then remove this interference. After interference cancellation, the static receiver has an equivalent point-to-point MIMO channel with perfect CSIR, in which case Gaussian signal achieves capacity.

Assuming $\mathbf{X}_2 \in \mathcal{X}_2$ has i.i.d. $\mathcal{CN}(0, 1)$ entries, the equivalent channel for the dynamic receiver $\mathbf{H}_1\mathbf{X}_2$ is isotropically

distributed (see Definition 1), which leads to two properties. First, for any $T \times T$ unitary matrix Φ ,

$$p(\mathbf{Y}_1\Phi | \mathbf{X}_1\Phi) = p(\mathbf{Y}_1 | \mathbf{X}_1). \quad (52)$$

Second, for any $N_1 \times N_1$ unitary matrix Ψ

$$p(\mathbf{Y}_1 | \Psi\mathbf{X}_1) = p(\mathbf{Y}_1 | \mathbf{X}_1). \quad (53)$$

Based on these properties, the optimal signaling structure for the channel of the dynamic receiver is a diagonal matrix⁵ times a unitary matrix [8], [7]. Therefore, choosing \mathcal{X}_1 to be isotropically distributed unitary matrices is not far from optimal.

D. Degrees of Freedom Region

In this section, we show that the Grassmannian-Euclidean superposition achieves the optimal DoF region under certain channel conditions.

1) $N_1 \leq N_2$: In this case, the optimal DoF region is as follows.

Corollary 6 ($N_1 \leq N_2$): When an M -antenna transmitter transmits to a dynamic receiver and a static receiver with N_1 and N_2 antennas, respectively, with the dynamic channel coherence time T , the DoF region is:

$$\begin{cases} d_1 \leq N_1(1 - \frac{N_1}{T}) \\ \frac{d_1}{N_1} + \frac{d_2}{N_2} \leq 1 \end{cases}. \quad (54)$$

Proof: An outer bound can be found when both receivers have CSIR. The DoF region of the coherent upper bound is [2]

$$\frac{d_1}{N_1} + \frac{d_2}{N_2} \leq 1. \quad (55)$$

An inner bound is attained by Grassmannian-Euclidean superposition, which reaches the boundary of (55) except for $d_1 > N_1(1 - N_1/T)$. However, the DoF of the dynamic receiver can never exceed $N_1(1 - N_1/T)$ (see Section II). Therefore, Grassmannian-Euclidean superposition achieves the DoF region. \blacksquare

2) $N_1 > N_2$: In this case, the Grassmannian-Euclidean superposition does not match the coherent outer bound (55), however, we can partially characterize the DoF region for broadcasting with degraded message sets [6] and in the case of the more capable channel [5]. For both cases the capacity region is characterized by:

$$\begin{cases} R_1 \leq I(\mathbf{U}; \mathbf{Y}_1) \\ R_1 + R_2 \leq I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{U}) + I(\mathbf{U}; \mathbf{Y}_1) \\ R_1 + R_2 \leq I(\mathbf{X}_2; \mathbf{Y}_2) \end{cases}, \quad (56)$$

where \mathbf{U} is an auxiliary random variable. From the last inequality we have

$$R_1 + R_2 \leq N_2 \log \rho + O(1), \quad (57)$$

that is

$$d_1 + d_2 \leq N_2. \quad (58)$$

⁵When the channel is i.i.d. Rayleigh fading this diagonal matrix should be identity at high SNR [7]. However, it remains unknown whether the optimal choice is an identity matrix at arbitrary SNR.

When $N_2 \geq (1 - N_1/T)N_1$, the inner bound in Corollary 4 coincides with the outer bound (58) for $0 \leq d_1 \leq N_2(1 - N_2/T)$, therefore, the DoF is established for this range. For $d_1 > N_2(1 - N_2/T)$, the inner and outer bounds do not match, but the gap is small when N_2 is close to N_1 .

When $N_2 < (1 - N_1/T)N_1$, the inner bound in Corollary 4 is inferior to orthogonal transmission and the problem remains open.

V. CONCLUSION

Signal superposition based on a multiplicative structure was proposed to improve the degrees of freedom of the MIMO broadcast channels when one receiver has full CSIR while the other receiver has no CSIR. Two superposition signaling methods were proposed, both based on product superposition. In the Grassmannian superposition, the transmit signal is a product of two Grassmannian codewords, producing higher DoF than orthogonal transmission while reception is still interference-free at both receivers. The Grassmannian-Euclidean superposition uses coherent signaling for the receiver with CSIR, and Grassmannian signaling for the receiver without CSIR. The latter method is shown to attain the optimal DoF region under a broad set of channel conditions.

It is possible to extend the results of this paper to more than two receivers. The set of receivers can be divided into two sets, one with and one without CSIR. At each point in time, the transmitter uses product superposition to broadcast to two users, one from each group. A scheduler selects the pair of users that is serviced at each time. The time-sharing parameters defining the overall rate region are as follows: one parameter determines how long a given pair is serviced (time sharing between pairs) and for each pair a parameter determines the operating point of the degree-of-freedom region of that pair. To facilitate the case where there are unequal number of dynamic and static users, the pair memberships are allowed to be non-unique, i.e., there may be two or more pairs that contain a given receiver. The overall rate region is the convex hull of all rate vectors corresponding to all values of the time-sharing parameters mentioned above.

APPENDIX I PROOF OF THEOREM 1

A. Achievable Rate for the Dynamic Receiver

The normalized received signal $\mathbf{Y}_1 \in \mathcal{C}^{N_1 \times T}$ at the dynamic receiver is

$$\mathbf{Y}_1 = \sqrt{\frac{T}{N_1}} \mathbf{H}_1 \mathbf{X}_2 \mathbf{X}_1 + \frac{1}{\sqrt{\rho}} \mathbf{W}_1, \quad (59)$$

where $\mathbf{H}_1 \in \mathcal{C}^{N_1 \times N_2}$ is the dynamic channel, $\mathbf{X}_1 \in \mathcal{C}^{N_1 \times T}$ and $\mathbf{X}_2 \in \mathcal{C}^{N_2 \times N_1}$ are the isotropically distributed, unitary signals for the dynamic and static receivers, respectively, and $\mathbf{W}_1 \in \mathcal{C}^{N_1 \times T}$ is additive Gaussian noise.

Let $\tilde{\mathbf{H}}_1 \triangleq \mathbf{H}_1 \mathbf{X}_2$ be the $N_1 \times N_1$ equivalent channel, and rewrite (59) as

$$\mathbf{Y}_1 = \sqrt{\frac{T}{N_1}} \tilde{\mathbf{H}}_1 \mathbf{X}_1 + \frac{1}{\sqrt{\rho}} \mathbf{W}_1. \quad (60)$$

The elements in $\tilde{\mathbf{H}}_1$ are

$$\tilde{h}_{ij} = [\tilde{\mathbf{H}}_1]_{i,j} = \sum_{k=1}^{N_2} h_{ik} x_{kj}, \quad 1 \leq i, j \leq N_1, \quad (61)$$

where $h_{ik} = [\mathbf{H}_1]_{ik}$ and $x_{kj} = [\mathbf{X}_2]_{kj}$. Note that h_{ik} is i.i.d. random variable with zero mean and unit variance, therefore,

$$\mathbb{E}[\tilde{h}_{ij}^* \tilde{h}_{mn}] = 0 \quad (i, j) \neq (m, n). \quad (62)$$

For $(i, j) = (m, n)$ we have

$$\mathbb{E}[|\tilde{h}_{ij}|^2] = \sum_{k=1}^{N_2} \mathbb{E}[|h_{ik}|^2 |x_{kj}|^2] \quad (63)$$

$$= \mathbb{E}\left[\sum_{k=1}^{N_2} |x_{kj}|^2\right] = 1, \quad (64)$$

where (64) holds because $\mathbb{E}[|h_{ik}|^2] = 1$ and each column of \mathbf{X}_2 has unit norm. Therefore, the equivalent channel $\tilde{\mathbf{H}}_1$ has uncorrelated entries with zero mean and unit variance.

We now find a lower bound for the mutual information

$$I(\mathbf{X}_1; \mathbf{Y}_1) = h(\mathbf{Y}_1) - h(\mathbf{Y}_1 | \mathbf{X}_1), \quad (65)$$

i.e., an achievable rate for the dynamic receiver. First, we find an upper bound for $h(\mathbf{Y}_1 | \mathbf{X}_1)$. Let \mathbf{y}_{1i} be the row i of \mathbf{Y}_1 . Using the independence bound on entropy:

$$h(\mathbf{Y}_1 | \mathbf{X}_1) \leq \sum_{i=1}^{N_1} h(\mathbf{y}_{1i} | \mathbf{X}_1). \quad (66)$$

Let $\tilde{\mathbf{h}}_i$ be the row i of $\tilde{\mathbf{H}}_1$. Then, conditioned on \mathbf{X}_1 the covariance of \mathbf{y}_{1i} is

$$\mathbb{E}[\mathbf{y}_{1i}^\dagger \mathbf{y}_{1i} | \mathbf{X}_1] = \frac{T}{N_1} \mathbf{X}_1^\dagger \mathbb{E}[\tilde{\mathbf{h}}_i^\dagger \tilde{\mathbf{h}}_i] \mathbf{X}_1 + \frac{1}{\rho} \mathbf{I}_T \quad (67)$$

$$= \frac{T}{N_1} \mathbf{X}_1^\dagger \mathbf{X}_1 + \frac{1}{\rho} \mathbf{I}_T, \quad (68)$$

where the last equality holds since all the elements in $\tilde{\mathbf{H}}_1$ are uncorrelated with zero mean and unit variance. In addition, given \mathbf{X}_1 , the vector \mathbf{y}_{1i} has zero mean, and therefore, $h(\mathbf{y}_{1i} | \mathbf{X}_1)$ is upper bounded by the differential entropy of a multivariate normal random vector with the same covariance [4]:

$$h(\mathbf{y}_{1i} | \mathbf{X}_1) \leq \log \det\left(\frac{T}{N_1} \mathbf{X}_1^\dagger \mathbf{X}_1 + \frac{1}{\rho} \mathbf{I}\right) \quad (69)$$

$$\leq N_1 \log\left(\frac{T}{N_1} + \frac{1}{\rho}\right) - (T - N_1) \log \rho. \quad (70)$$

Combining (66) and (70), we obtain

$$h(\mathbf{Y}_1 | \mathbf{X}_1) \leq N_1^2 \log\left(\frac{T}{N_1} + \frac{1}{\rho}\right) - N_1(T - N_1) \log \rho. \quad (71)$$

After calculating the upper bound for $h(\mathbf{Y}_1 | \mathbf{X}_1)$, we now find a lower bound for $h(\mathbf{Y}_1)$ as follows.

$$h(\mathbf{Y}_1) > h\left(\sqrt{\frac{T}{N_1}} \mathbf{H}_1 \mathbf{X}_2 \mathbf{X}_1\right) \quad (72)$$

$$\geq h\left(\sqrt{\frac{T}{N_1}} \mathbf{H}_1 \mathbf{X}_2 \mathbf{X}_1 \mid \mathbf{H}_1, \mathbf{X}_2\right), \quad (73)$$

where (72) holds since we remove the noise, and (73) holds since conditioning does not increase differential entropy. The Jacobian from \mathbf{X}_1 to $\mathbf{H}_1\mathbf{X}_2\mathbf{X}_1$ is [15, Theorem. 2.1.5]:

$$J_{X_1} = \left(\sqrt{\frac{T}{N_1 N_2}} \det(\mathbf{H}_1 \mathbf{X}_2) \right)^{N_1}. \quad (74)$$

Therefore, from (73) we have

$$h(\mathbf{Y}_1) > h(\mathbf{X}_1) + \mathbb{E}[\log J_{X_1}], \quad (75)$$

where the expectation is with respect to \mathbf{X}_2 and \mathbf{H}_1 . Because \mathbf{X}_1 is an isotropically distributed unitary matrix, i.e., uniformly distributed on the Stiefel manifold $\mathbb{F}(T, N_1)$, we have [7]

$$h(\mathbf{X}_1) = \log |\mathbb{F}(T, N_1)|, \quad (76)$$

where $|\mathbb{F}(T, N_1)|$ is the volume of $\mathbb{F}(T, N_1)$ based on the Haar measure induced by the Lebesgue measure restricted to the Stiefel manifold [15]:

$$|\mathbb{F}(T, N_1)| = \prod_{i=T-N_1+1}^T \frac{2\pi^i}{(i-1)!}. \quad (77)$$

Finally, combining (71) and (75), we obtain

$$\begin{aligned} I(\mathbf{X}_1; \mathbf{Y}_1) &> N_1(T - N_1) \log \rho + \log |\mathbb{F}(T, N_1)| \\ &\quad + \mathbb{E}[\log J_{X_1}] - N_1 \sum_{i=1}^{N_1} \log \left(\frac{T}{N_1} + \frac{1}{\rho} \right) \end{aligned} \quad (78)$$

$$= N_1(T - N_1) \log \rho + O(1). \quad (79)$$

Normalizing $I(\mathbf{X}_1; \mathbf{Y}_1)$ over T time-slots yields the achievable rate of the dynamic receiver.

B. Achievable Rate for the Static Receiver

The signal received at the static receiver is

$$\mathbf{Y}_2 = \sqrt{\frac{T}{N_1}} \mathbf{H}_2 \mathbf{X}_2 \mathbf{X}_1 + \frac{1}{\sqrt{\rho}} \mathbf{W}_2, \quad (80)$$

where $\mathbf{H}_2 \in \mathcal{C}^{N_2 \times N_2}$ is the static channel and $\mathbf{W}_2 \in \mathcal{C}^{N_2 \times T}$ is additive Gaussian noise. Denote the sub-matrix containing the first N_1 columns of \mathbf{Y}_2 with \mathbf{Y}'_2 .

$$\mathbf{Y}'_2 = \sqrt{\frac{T}{N_1}} \mathbf{H}_2 \mathbf{X}_2 \mathbf{X}'_1 + \frac{1}{\sqrt{\rho}} \mathbf{W}'_2, \quad (81)$$

where $\mathbf{X}'_1 \in \mathcal{C}^{N_1 \times N_1}$ is the corresponding sub-matrix of \mathbf{X}_1 , and $\mathbf{W}'_2 \in \mathcal{C}^{N_1 \times N_1}$ is i.i.d. Gaussian noise. Given \mathbf{H}_2 , the mutual information between \mathbf{Y}_2 and \mathbf{X}_2 is lower bounded by:

$$I(\mathbf{Y}_2; \mathbf{X}_2 | \mathbf{H}_2) \geq I(\mathbf{Y}'_2; \mathbf{X}_2 | \mathbf{H}_2). \quad (82)$$

We will focus on $I(\mathbf{Y}'_2; \mathbf{X}_2 | \mathbf{H}_2)$ to derive a lower bound. Using the singular value decomposition (SVD):

$$\mathbf{H}_2 = \mathbf{U}_2^\dagger \mathbf{\Sigma}_2 \mathbf{V}_2, \quad (83)$$

where $\mathbf{U}_2, \mathbf{V}_2 \in \mathcal{C}^{N_2 \times N_2}$ and $\mathbf{\Sigma}_2 = \text{diag}(\lambda_1, \dots, \lambda_{N_2})$ with $|\lambda_1| \geq \dots \geq |\lambda_{N_2}|$. Since \mathbf{H}_2 is known and non-singular, the dynamic receiver applies \mathbf{H}_2^{-1} to remove it:

$$\mathbf{H}_2^{-1} \mathbf{Y}'_2 = \sqrt{\frac{T}{N_1}} \mathbf{X}_2 \mathbf{X}'_1 + \frac{1}{\sqrt{\rho}} \mathbf{W}_2''. \quad (84)$$

The columns of \mathbf{W}_2'' are mutually independent, and each column has an autocorrelation:

$$\mathbf{R}_W = \mathbf{V}_2^\dagger \mathbf{\Sigma}_2^{-2} \mathbf{V}_2. \quad (85)$$

Because mutual information is independent of the choice of coordinates, we have

$$I(\mathbf{Y}'_2; \mathbf{X}_2 | \mathbf{H}_2) = I(\mathbf{H}_2^{-1} \mathbf{Y}'_2; \mathbf{X}_2 | \mathbf{H}_2) \quad (86)$$

$$= h(\mathbf{H}_2^{-1} \mathbf{Y}'_2 | \mathbf{H}_2) - h(\mathbf{H}_2^{-1} \mathbf{Y}'_2 | \mathbf{X}_2, \mathbf{H}_2). \quad (87)$$

Let $\mathbf{y}_{2,i}$ be the column i of $\mathbf{H}_2^{-1} \mathbf{Y}'_2$, then via the independence bound on entropy:

$$h(\mathbf{H}_2^{-1} \mathbf{Y}'_2 | \mathbf{X}_2, \mathbf{H}_2) \leq \sum_{i=1}^{N_1} h(\mathbf{y}_{2,i} | \mathbf{X}_2, \mathbf{H}_2). \quad (88)$$

From (84) and (85), the autocorrelation of $\mathbf{y}_{2,i}$ conditioned on \mathbf{X}_2 and \mathbf{H}_2 is

$$\mathbf{R}_{2,i} = \frac{T}{N_1} \mathbf{X}_2 \mathbb{E}[\mathbf{x}'_{1,i} \mathbf{x}_{1,i}^\dagger] \mathbf{X}_2^\dagger + \frac{1}{\rho} \mathbf{R}_W \quad (89)$$

$$= \frac{T}{N_1} \mathbf{X}_2 \mathbf{R}_{1,i} \mathbf{X}_2^\dagger + \frac{1}{\rho} \mathbf{R}_W \quad (90)$$

where $\mathbf{x}'_{1,i} \in \mathcal{C}^{N_1 \times 1}$ is the column i of \mathbf{X}'_1 and has autocorrelation $\mathbf{R}_{1,i}$. The expected value of $\mathbf{y}_{2,i}$ is zero and thus the differential entropy is maximized if $\mathbf{y}_{2,i}$ has multivariate normal distribution [4]:

$$\begin{aligned} h(\mathbf{y}_{2,i} | \mathbf{X}_2, \mathbf{H}_2) &\leq \log \det \left(\frac{T}{N_1} \mathbf{X}_2 \mathbf{R}_{1,i} \mathbf{X}_2^\dagger + \frac{1}{\rho} \mathbf{R}_W \right) \\ &= \log \det \left(\frac{T}{N_1} \mathbf{V}_2 \mathbf{X}_2 \mathbf{R}_{1,i} \mathbf{X}_2^\dagger \mathbf{V}_2^\dagger + \frac{1}{\rho} \mathbf{\Sigma}_2^{-2} \right). \end{aligned} \quad (91)$$

The following lemma calculates $\mathbf{R}_{1,i}$, the autocorrelation of a column of an i.d. matrix.

Lemma 1: If $\mathbf{Q} \in \mathcal{C}^{T \times T}$ is isotropically distributed (i.d.) unitary matrix, then each row and column of \mathbf{Q} is an i.d. unit vector with autocorrelation $\frac{1}{T} \mathbf{I}_T$.

Proof: From Definition 1, given \mathbf{Q} is i.d., for any deterministic unitary matrix $\Phi \in \mathcal{C}^{T \times T}$,

$$p(\mathbf{Q}\Phi) = p(\mathbf{Q}), \quad (92)$$

which implies that the marginal distribution of each row and column remains unchanged under any transform Φ . Therefore, each row and column is an i.d. unit vector. Without loss of generality, we consider the first row of \mathbf{Q} , denoted as \mathbf{q}_1 . Let the autocorrelation of \mathbf{q}_1 be \mathbf{R}_q and posses the eigenvalue decomposition $\mathbf{R}_q = \mathbf{P}^\dagger \mathbf{\Sigma}_q \mathbf{P}$, where $\mathbf{P} \in \mathcal{C}^{n \times n}$ is unitary and $\mathbf{\Sigma}_q$ is diagonal. Since $\mathbf{q}_1 \mathbf{P}^\dagger$ has the same distribution as \mathbf{q}_1 , therefore

$$\mathbf{R}_q = \mathbb{E}[\mathbf{q}_1^\dagger \mathbf{q}_1] = \mathbf{P} \mathbb{E}[\mathbf{q}_1^\dagger \mathbf{q}_1] \mathbf{P}^\dagger = \mathbf{\Sigma}_q. \quad (93)$$

Thus \mathbf{R}_q is a diagonal matrix. Furthermore, the diagonal elements of $\mathbf{\Sigma}_q$ have to be identical, i.e., $\mathbf{\Sigma}_q = a \mathbf{I}_T$, otherwise \mathbf{R}_q would not be rotationally invariant which conflicts with the i.d. assumption. Finally, because $\text{tr}(\mathbf{R}_q) = 1$, we have $\mathbf{\Sigma}_q = \mathbf{I}_T/T$. This completes the proof of Lemma 1. ■

Since \mathbf{X}_1 is an i.d. unitary matrix, based on Lemma 1, the autocorrelation of its sub-column is

$$\mathbf{R}_{1,i} = \mathbf{I}_{N_1}/T. \quad (94)$$

Therefore, the eigenvalues of $\mathbf{V}_2 \mathbf{X}_2 \mathbf{R}_{1,i} \mathbf{X}_2^\dagger \mathbf{V}_2^\dagger$ are

$$\left(\underbrace{\frac{1}{T}, \dots, \frac{1}{T}}_{N_1}, \underbrace{0, \dots, 0}_{N_2 - N_1} \right). \quad (95)$$

We now bound the eigenvalues of the sum of two matrices in (91), noting that λ_j^{-2} are in ascending order and using a theorem of Weyl [16, Theorem 4.3.1]:

$$\begin{aligned} h(\mathbf{y}_{2,i} | \mathbf{X}_2, \mathbf{H}_2) &\leq N_1 \log \left(\frac{1}{N_1} + \lambda_{N_2}^{-2} \right) \\ &\quad + (N_2 - N_1) \log \frac{1}{\rho} \lambda_{N_2}^{-2}. \end{aligned} \quad (96)$$

From (88) and (96), we have:

$$\begin{aligned} h(\mathbf{H}_2^{-1} \mathbf{Y}_2' | \mathbf{X}_2, \mathbf{H}_2) &\leq N_1^2 \log \left(\frac{1}{N_1} + \lambda_{N_2}^{-2} \right) \\ &\quad - N_1(N_2 - N_1) \log \lambda_{N_2}^{-2} - N_1(N_2 - N_1) \log \rho. \end{aligned} \quad (97)$$

We now calculate a lower bound for $h(\mathbf{H}_2^{-1} \mathbf{Y}_2' | \mathbf{H}_2)$:

$$h(\mathbf{H}_2^{-1} \mathbf{Y}_2' | \mathbf{H}_2) > h\left(\sqrt{\frac{T}{N_1}} \mathbf{X}_2 \mathbf{X}_1' | \mathbf{H}_2\right) \quad (98)$$

$$> h\left(\sqrt{\frac{T}{N_1}} \mathbf{X}_2 \mathbf{X}_1' | \mathbf{X}_1', \mathbf{H}_2\right). \quad (99)$$

From [15, Theorem. 2.1.5], given \mathbf{X}_1' the Jacobian of the transformation from \mathbf{X}_2 to $\sqrt{\frac{T}{N_1}} \mathbf{X}_2 \mathbf{X}_1'$ is:

$$J_{X_2} = \left(\sqrt{\frac{T}{N_1}} \right)^{N_2} \det(\mathbf{X}_1')^{N_1}. \quad (100)$$

Therefore, from the right hand side of (99) we have

$$h(\mathbf{H}_2^{-1} \mathbf{Y}_2' | \mathbf{H}_2) > h(\mathbf{X}_2) + \mathbb{E}[\log J_{X_2}], \quad (101)$$

where the expectation is with respect to \mathbf{X}_1' . Because \mathbf{X}_2 is uniformly distributed on the Stiefel manifold $\mathbb{F}(N_2, N_1)$, we have [7]

$$h(\mathbf{X}_1) = \log |\mathbb{F}(N_2, N_1)|, \quad (102)$$

where $|\mathbb{F}(N_2, N_1)|$ is the volume of $\mathbb{F}(N_2, N_1)$, which is given by [15]:

$$|\mathbb{F}(N_2, N_1)| = \prod_{i=N_2-N_1+1}^{N_2} \frac{2\pi^i}{(i-1)!}. \quad (103)$$

Finally, substituting (101) and (97) into (87), we have

$$I(\mathbf{Y}_2'; \mathbf{X}_2 | \mathbf{H}_2) = N_1(N_2 - N_1) \log \rho + O(1). \quad (104)$$

Hence, the rate achieved by the static receiver is

$$\frac{1}{T} \mathbb{E}[I(\mathbf{Y}_2'; \mathbf{X}_2 | \mathbf{H}_2)] = \frac{N_1}{T} (N_2 - N_1) \log \rho + O(1), \quad (105)$$

where the expectation is with respect to \mathbf{H}_2 .

APPENDIX II PROOF OF COROLLARY 1

The objective is to find the best dimensions for the transmit signals $\mathbf{X}_1 \in \mathcal{C}^{\hat{N}_1 \times \hat{T}}$ and $\mathbf{X}_2 \in \mathcal{C}^{\hat{N}_2 \times \hat{N}_1}$. From Theorem 1, it is easily determined that $\hat{N}_2 = N_2$ is optimal, because the pre-log factor of R_2 increases with \hat{N}_2 and the pre-log factor of R_1 is independent of \hat{N}_2 (given $\hat{N}_1 \leq N_2$).

To find the optimal values of \hat{N}_1, \hat{T} , we start by relaxing the variables by allowing them to be continuous valued, i.e. $\hat{N}_1 \rightarrow x$ and $\hat{T} \rightarrow y$, and then showing via the derivatives that the cost functions are monotonic, therefore optimal values reside at the boundaries, which are indeed integers.

Using the DoF expression from Theorem 1, the slope between two achievable points \mathcal{D}_2 and \mathcal{D}_3 is:

$$f(x, y) = \frac{x(N_2 - x)/y - N_2}{x(1 - x/y)}. \quad (106)$$

Therefore, for all $0 < x \leq N_1$,

$$\frac{\partial f(x, y)}{\partial y} = \frac{x}{(y - x)^2} > 0. \quad (107)$$

We wish to maximize f with the constraint $y \leq T$, thus $y = T$ is optimal.

Substituting $y = T$ into $f(x, y)$, we have

$$\frac{\partial f(x, T)}{\partial x} = -\frac{(T - N_2)x^2 + TN_2x - T^2N_2}{x^2(T - x)^2}. \quad (108)$$

If $T = N_2$, since $x \leq T/2$, then $\frac{\partial f}{\partial x} > 0$. In this case $x = N_1$ maximizes the DoF region.

If $T \neq N_2$, let $T = \alpha N_2$. When $0 < \alpha < \frac{3}{4}$, one can verify that $\frac{\partial f}{\partial x} > 0$ for all $x > 0$. Thus, $x = N_1$ is optimal. When $\alpha \geq \frac{3}{4}$, let $\frac{\partial f}{\partial x} = 0$, and we have the corresponding solutions:

$$x_{1,2} = \frac{-\alpha N_2 \pm \alpha N_2 \sqrt{1 + 4(\alpha - 1)}}{2(\alpha - 1)}. \quad (109)$$

When $\frac{3}{4} \leq \alpha < 1$, the above solutions are positive, where the smaller one is:

$$x_1 = \frac{\alpha N_2 - \alpha N_2 \sqrt{1 + 4(1 - \alpha)}}{2(1 - \alpha)} > N_1. \quad (110)$$

Since $\frac{\partial f}{\partial x} > 0$ at $x = 0$, we have $\frac{\partial f}{\partial x} > 0$ for $0 \leq x \leq N_1$. When $\alpha > 1$, the (only) positive solution of (109) is:

$$x_1 = \frac{\alpha N_2 + \alpha N_2 \sqrt{1 + 4(\alpha - 1)}}{2(\alpha - 1)} > N_1. \quad (111)$$

Once again, since $\frac{\partial f}{\partial x} > 0$ at $x = 0$, we have $\frac{\partial f}{\partial x} > 0$ for $0 \leq x \leq N_1$.

Therefore, for all cases, $x = N_1$ maximizes the DoF region.

APPENDIX III PROOF OF THEOREM 2

A. Achievable Rate for the Dynamic Receiver

The proof is similar to the proof for Theorem 1, so we only outline key steps. The received signal at the dynamic receiver is

$$\mathbf{Y}_1 = \sqrt{\frac{T}{N_1 N_2}} \mathbf{H}_1 \mathbf{X}_2 \mathbf{X}_1 + \frac{1}{\sqrt{\rho}} \mathbf{W}_1, \quad (112)$$

where $\mathbf{Y}_1 \in \mathcal{C}^{N_1 \times T}$ and $\mathbf{H}_1 \in \mathcal{C}^{N_1 \times N_2}$ and $\mathbf{W}_1 \in \mathcal{C}^{N_1 \times T}$ is additive Gaussian noise. We establish a lower bound for the mutual information between \mathbf{X}_1 and \mathbf{Y}_1 :

$$I(\mathbf{X}_1; \mathbf{Y}_1) = h(\mathbf{Y}_1) - h(\mathbf{Y}_1 | \mathbf{X}_1). \quad (113)$$

In the above equation, we have

$$h(\mathbf{Y}_1 | \mathbf{X}_1) \leq \sum_{i=1}^{N_1} h(\mathbf{y}_{1i} | \mathbf{X}_1).$$

One can verify

$$h(\mathbf{y}_{1i} | \mathbf{X}_1) \leq \log \det \left(\frac{T}{N_2} \mathbf{X}_1^\dagger \mathbf{X}_1 + \frac{1}{\rho} \mathbf{I} \right). \quad (114)$$

Finally, we obtain

$$h(\mathbf{Y}_1 | \mathbf{X}_1) < N_1^2 \log \left(\frac{T}{N_2} + \frac{1}{\rho} \right) - N_1(T - N_1) \log \rho. \quad (115)$$

The lower bound is given by:

$$h(\mathbf{Y}_1) > \log |\mathbb{F}(T, N_1)| + \mathbb{E}[\log J_{X_1}], \quad (116)$$

where the expectation is with respect to \mathbf{H}_1 and \mathbf{X}_2 , and

$$J_{X_1} = \left(\sqrt{\frac{T}{N_1 N_2}} \det(\mathbf{H}_1 \mathbf{X}_2) \right)^{N_1}. \quad (117)$$

Combining (115) and (117), and normalizing over T time-slots leads to the achievable rate of the dynamic receiver.

B. Achievable Rate for the Static Receiver

The received signal at the static receiver is $\mathbf{Y}_2 \in \mathcal{C}^{N_2 \times T}$

$$\mathbf{Y}_2 = \sqrt{\frac{T}{N_1 N_2}} \mathbf{H}_2 \mathbf{X}_2 \mathbf{X}_1 + \frac{1}{\sqrt{\rho}} \mathbf{W}_2,$$

where $\mathbf{H}_2 \in \mathcal{C}^{N_2 \times N_2}$ is the static channel, and $\mathbf{W}_2 \in \mathcal{C}^{N_2 \times T}$ is additive Gaussian noise.

We first calculate the *decodable* dynamic rate at the static receiver in the next lemma.

Lemma 2: The static receiver is able to decode the dynamic rate R_1 if

$$R_1 \leq N_1(1 - N_1/T) \log \rho + O(1). \quad (118)$$

Proof: Use the SVD for \mathbf{H}_2 and re-write the signal at the static receiver as

$$\mathbf{Y}_2 = \sqrt{\frac{T}{N_1 N_2}} \mathbf{U}_2^\dagger \Sigma_2 \mathbf{V}_2 \mathbf{X}_2 \mathbf{X}_1 + \frac{1}{\sqrt{\rho}} \mathbf{W}_2, \quad (119)$$

Because \mathbf{X}_2 is an isotropically distributed unitary matrix, $\mathbf{X}_2' \triangleq \mathbf{V}_2 \mathbf{X}_2$ has the same distribution as \mathbf{X}_2 , i.e., a matrix of i.i.d. $\mathcal{CN}(0, 1)$. Rotate \mathbf{Y}_2 with \mathbf{U}_2

$$\mathbf{Y}_2' \triangleq \mathbf{U}_2 \mathbf{Y}_2 = \sqrt{\frac{T}{N_1 N_2}} \Sigma_2 \mathbf{X}_2' \mathbf{X}_1 + \frac{1}{\sqrt{\rho}} \mathbf{W}_2', \quad (120)$$

where \mathbf{W}_2' is i.i.d. Gaussian noise. Let $\mathbf{Y}_2'' \in \mathcal{C}^{N_1 \times T}$ be the first N_1 rows of \mathbf{Y}_2' , i.e., the rows corresponding to the largest N_1 singular modes of \mathbf{H}_2 , that is $|\lambda_1| \geq \dots \geq |\lambda_{N_1}|$. We denote the corresponding $N_1 \times N_1$ sub-matrix of \mathbf{X}_2' by \mathbf{X}_2'' . Then,

$$\mathbf{Y}_2'' = \text{diag}(\lambda_1, \dots, \lambda_{N_1}) \mathbf{X}_2'' \mathbf{X}_1 + \frac{1}{\sqrt{\rho}} \mathbf{W}_2''. \quad (121)$$

Conditioned on \mathbf{H}_2 , the decodable dynamic rate at the static receiver is

$$I(\mathbf{X}_1; \mathbf{Y}_2 | \mathbf{H}_2) = I(\mathbf{X}_1; \mathbf{Y}_2' | \mathbf{H}_2),$$

which is lower bounded by

$$I(\mathbf{X}_1; \mathbf{Y}_2'' | \mathbf{H}_2) = h(\mathbf{Y}_2'' | \mathbf{H}_2) - h(\mathbf{Y}_2'' | \mathbf{X}_1, \mathbf{H}_2). \quad (122)$$

Using the independence bound for $h(\mathbf{Y}_2'' | \mathbf{X}_1, \mathbf{H}_2)$ yields

$$h(\mathbf{Y}_2'' | \mathbf{X}_1, \mathbf{H}_2) \leq \sum_{i=1}^{N_1} h(\mathbf{y}_{2i} | \mathbf{X}_1, \mathbf{H}_2), \quad (123)$$

where \mathbf{y}_{2i} is the row i of \mathbf{Y}_2'' . Let \mathbf{x}_{2i} be the row i of \mathbf{X}_2'' , for $1 \leq i \leq N_1$. Since $\mathbf{X}_2'' \in \mathcal{C}^{N_1 \times N_1}$ have i.i.d. $\mathcal{CN}(0, 1)$ entries, all the row vectors \mathbf{x}_{2i} have the same autocorrelation I_{N_1} .

Conditioned on \mathbf{X}_1 , the autocorrelation of $\mathbf{y}_{2i} = \lambda_i \mathbf{x}_{2i} \mathbf{X}_1$ is given by

$$\mathbb{E}[\mathbf{y}_{2i}^\dagger \mathbf{y}_{2i} | \mathbf{X}_1, \mathbf{H}_2] = \lambda_i^2 \mathbf{X}_1^\dagger \mathbf{X}_1 + \frac{1}{\rho} \mathbf{I}_T. \quad (124)$$

Therefore,

$$h(\mathbf{y}_{2i} | \mathbf{X}_1, \mathbf{H}_2) \leq \log \det \left(\lambda_i^2 \mathbf{X}_1^\dagger \mathbf{X}_1 + \frac{1}{\rho} \mathbf{I}_T \right), \quad (125)$$

$$= N_1 \log \left(\lambda_i^2 + \frac{1}{\rho} \right) - (T - N_1) \log \rho. \quad (126)$$

and subsequently,

$$h(\mathbf{Y}_2'' | \mathbf{X}_1, \mathbf{H}_2) \leq \sum_{i=1}^{N_1} \log \left(\lambda_i^2 + \frac{1}{\rho} \right) - N_1(T - N_1) \log \rho. \quad (127)$$

We now find a lower bound for $h(\mathbf{Y}_2'' | \mathbf{H}_2)$. Similar to (101), we have

$$h(\mathbf{Y}_2'' | \mathbf{H}_2) \geq h(\mathbf{X}_1) + \mathbb{E}[J_{X_2}], \quad (128)$$

where the expectation is with respect to \mathbf{X}_2 , and

$$h(\mathbf{X}_1) = |\mathbb{F}(T, N_1)| = \prod_{i=T-N_1+1}^T \frac{2\pi^i}{(i-1)!}, \quad (129)$$

and

$$J_{X_2} = \prod_{i=1}^{N_1} \lambda^{2N_1} \det(\mathbf{X}_2)^{N_1}. \quad (130)$$

Finally, taking expectation over \mathbf{H}_2 , we obtain

$$\begin{aligned} \mathbb{E}[I(\mathbf{X}_1; \mathbf{Y}_2 | \mathbf{H}_2)] &\geq N_1(T - N_1) \log \rho + h(\mathbf{X}_1) + \mathbb{E}[J_{X_2}] \\ &\quad - \mathbb{E} \left[\sum_{i=1}^{N_1} \log \left(\lambda_i^2 + \frac{1}{\rho} \right) \right] \\ &= N_1(T - N_1) \log \rho + O(1). \end{aligned} \quad (131)$$

This completes the proof for Lemma 2. \blacksquare

Therefore, the transmitter is able to send $N_1(1 - N_1/T)$ DoF to the dynamic receiver, while ensuring the dynamic signal is decoded at the static receiver.

After decoding \mathbf{X}_1 , the static receiver removes the interference:

$$\mathbf{Y}_2 \mathbf{X}_1^\dagger = \sqrt{\frac{T}{N_1 N_2}} \mathbf{H}_2 \mathbf{X}_2 + \frac{1}{\sqrt{\rho}} \mathbf{W}_2', \quad (132)$$

where $\mathbf{W}'_2 \in \mathcal{C}^{N_2 \times N_1}$ is the equivalent noise whose entries are still i.i.d. $\mathcal{CN}(0, 1)$. The equivalent channel for the static receiver is now a point-to-point MIMO channel. With Gaussian input \mathbf{X}_2 , we have [13]

$$I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{H}_2) = N_1 N_2 \log \rho + O(1). \quad (133)$$

Normalizing $I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{H}_2)$ over T time-slots yields the achievable rate of the static receiver.

REFERENCES

- [1] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691 – 1706, July 2003.
- [2] C. Huang, S. Jafar, S. Shamai, and S. Vishwanath, "On degrees of freedom region of MIMO networks without channel state information at transmitters," pp. 849 – 857, Feb. 2012.
- [3] S. Jafar and A. Goldsmith, "Isotropic fading vector broadcast channels: the scalar upper bound and loss in degrees of freedom," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 848 – 857, Mar. 2005.
- [4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.
- [5] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inform. Theory*, vol. 25, no. 2, pp. 166 – 169, 1979.
- [6] J. Korner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. 23, no. 1, pp. 60 – 64, 1977.
- [7] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359–383, 2002.
- [8] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, 1999.
- [9] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 1986.
- [10] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2001.
- [11] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 543–564, 2000.
- [12] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, 2003.
- [13] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. on Telecomm.*, vol. 10, no. 6, pp. 585–595, 1999.
- [14] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthonormality constraints," *Appl. Comput. Harmonic Anal.*, vol. 20, no. 2, pp. 303–353, 1998.
- [15] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. Wiley, 1982.
- [16] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

Aria Nosratinia (S'87-M'97-SM'04-F'10) is Jonsson Distinguished Professor of Engineering at the University of Texas at Dallas. He received his Ph.D. in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 1996. He has held visiting appointments at Princeton University, Rice University, and UCLA. His interests lie in the broad area of information theory and signal processing, with applications in wireless communications. He was the secretary of the IEEE Information Theory Society in 2010-2011 and the treasurer for ISIT 2010 in Austin, Texas. He has served as editor for the IEEE Transactions on Information Theory, IEEE Transactions on Wireless Communications, IEEE Signal Processing Letters, IEEE Transactions on Image Processing, and IEEE Wireless Communications (Magazine). He has been the recipient of the National Science Foundation career award, and is a fellow of IEEE.

Yang Li (S'10) received his B.S. and M.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China in 2005 and 2008, respectively. He is currently pursuing the Ph.D degree in electrical engineering at the University of Texas at Dallas. He has interned at Samsung Telecommunications America in 2012, and at Huawei Technologies Co. Ltd in 2011 and 2008. His current interests include cognitive radio, heterogeneous network, interference management and cooperative communication.